How fluent is the fluent speech of people who stutter? A new approach to measuring kinematics with ultrasound
2016

Cornelia J. Heyde,James M. Scobbie,Robin Lickley & Eleanor K. E. Drake

https://doi.org/10.3109/02699206.2015.1100684

Clinical Linguistics & Phonetics

Volume 30, 2016 (Issues 3-5):

Insights from Ultrasound: Enhancing Our Understanding of Clinical Phonetics

Pages 292-312

Version of Record Received 15 Feb 2015, Accepted 22 Sep 2015, Published online: 23 Nov 2015

How fluent is the fluent speech of people who stutter? A new approach to measuring kinematics with ultrasound

Abstract

We present a new approach to the investigation of dynamic ultrasound tongue imaging (UTI) data, applied here to analyse subtle aspects of the fluency of people who stutter (PWS). Fluent productions of CV syllables (C=/k/; V=/ɑ, i, ə/) from three PWS and three control speakers (PNS) were analysed for duration and peak velocity relative to articulatory movement towards (onset) and away from (offset) the consonantal closure. The objective was to apply a replicable methodology for kinematic investigation to speech of PNS in order to test Wingate's Fault-Line hypothesis. As was hypothesised, results show comparable onset behaviours for both groups. Regarding offsets, groups differ in peak velocity. Results suggest that PWS do not struggle initiating consonantal closure (onset). In transition from consonantal closure into the vowel, however, groups appear to employ different strategies expressed in increased variation (PNS) versus decreased mean peak velocity (PWS).

Title: How fluent is the fluent speech of people who stutter? A new approach to measuring kinematics with ultrasound

Introduction

Persistent developmental stuttering is a motor-speech disorder (Namasivayam & van Lieshout, 2011) which emerges in childhood. It is typically characterized by a relapsing-remitting, often situation-specific pattern of symptoms; primarily involuntary disruptions in the smooth flow of speech. These symptoms are described in terms of their acoustic consequences, labelled as blocks, prolongations and repetitions. The majority of the motor disruption underlying these acoustic consequences occurs within the (internal) vocal tract. It is therefore difficult to observe and measure the speech motor activity directly involved in stuttering. For the same reason it is usually difficult to compare the speech-motor performance during fluent speech of people who stutter and those who do not, which is an important task if we hope to understand the sources of the disruptions. Ultrasound tongue imaging (UTI) offers a means to observe the speech-motor activity of the primary active oral articulator. It therefore has much to offer the study of stuttering, particularly in light of the suggestion that stuttering is best understood as involving disruption to the high temporal coordination of oral (articulatory) and laryngeal (phonatory) movements (Van Riper, 1982; Adams, 1999; Max & Gracco, 2005 for a review). In this paper we report a methodology we have adopted for investigating the dynamics of articulatory motor-speech production, both in PWS and in PNS. We will provide descriptive findings comparing the speech-motor productions of 3 PWS to 3 PNS using this ultrasound-based analysis.

Under experimental conditions, PWS perform more poorly across a range of acoustic measures of speech performance than do PNS. By their own rating and that of others, PWS are more susceptible to speech error elicitation than are PNS (Brocklehurst & Corley, 2011). PWS as a group also have longer speech reaction times (Cross & Luper, 1979; Horii, 1984; Harbison, Porter, & Tobey,

1989). Group differences between PWS and PNS in voice onset times (VOT) may be observable only in specific phonetic or utterance contexts (Watson & Alfonso, 1982; Healey & Ramig, 1986; De Nil & Brutten, 1991). As a group, PWS have been found to have longer vowel and consonant durations than PNS (Di Simoni, 1974; Starkweather & Myers, 1979). 1979). PWS were found to have descriptively longer VOT than PNS (Bakker & Brutten, 1990).

Table 1. Studies investigating the speech and non-speech motor performance of people who stutter

Study	Population	Instrumental approach	Topic investigated	Key findings
Chang, Ohde, & Conture (2002)	Children who stutter (CWS) v. children who do not stutter (CNS)	Acoustic measurement of formant transitions and F2 for CV syllables	Place of articulation and formant transitions	Groups differ in formant transition rate (FTR) as a function of place of articulation. CWS exhibit less contrast of FTRs between the labial and alveolar consonant contexts than CNS
Namasivayam & van Lieshout (2008)	Adults who stutter (AWS) v. adults who do not stutter (ANS)	Electromagnetic articulography (EMA) Transducer coils on midline of vermilion border of upper and lower lips (UL, LL), lower jaw (J), the tongue blade (c. 1cm behind the anatomical tongue tip), the tongue body (c. 3cm behind tongue blade coil) and the tongue dorsum (c. 2cm behind tongue body coil). Only report bilabial productions.	Intergestural timing and stability.	Amplitude of UL movement was significantly larger in PWS than PNS across normal and fast speech rates
McClean, Tasko, & Runyan (2004)	AWS/ANS	EMA (UL, LL, TB, J)	Velocity, duration and speed ratios of different articulators	Complex pattern of findings: Task complexity interacted selectively with articulatory features
Smith, Sadagopan, Walsh, & Weber-Fox (2010)	AWS/ANS	Optotrak 3020 motion tracking system, tracking infrared light emitting diodes (IREDs) attached to the upper and lower lip (vermillion border). Tested nonword productions.	Articulatory stability	Higher lip aperture variability in AWS, especially in early trials compared to later trials.

Kleinow & Smith (2000)	AWS	IREDS attached to lower lip. Tested real words productions in carrier phrases	Articulatory stability	Greater variability in AWS, who were also vulnerable to the phonological complexity of words whereas ANS were not.
Caruso, Abbs & Gracco (1988)	AWS/ANS	Strain gauge on UL, LL,J.	Inter-articulator sequencing	Between-group differences in the sequencing of movement onsets and velocity peaks
Max, Caruso, & Gracco (2003)	AWS/ANS	Speech, non-speech and finger movements. Tested real nouns with bilabial onsets, following 'my'. Used UL, LL, jaw strain gauge.		Between-group difference on lip and jaw closing (but not opening). AWS showed both longer movement durations and higher peak velocities and greater amplitudes during closing movements
Max & Gracco (2005)	AWS/ANS	EMA and EGG UL, LL, J and larynx	Inter-articulator sequencing	Longer acoustic durations for voice onset time and devoicing intervals for AWS. Group differences in kinematics of oral and laryngeal gesture coordination as measured by onset and peak velocity and vocal fold vibration (i.e. AWS show longer duration between laryngeal and oral onsets of movement)
Zimmermann (1980)	AWS/ANS	Cineradiography LL and jaw	Inter-articulator sequencing	Longer transition times and longer steady-state postures for AWS. Movements of AWS show greater asynchrony than those of ANS

.

It is apparent that the poorer speech performance of PWS on acoustic measures reflects an underlying motor deficit of some nature. Between group differences have been found for both non-speech and speech oro-motor performance (cf. Table 1). However, the fluctuating severity of stuttering symptoms indicates that the nature of the underlying motor deficit is probably complex and subtle: PWS are capable of producing speech that is acoustically indistinguishable from the speech of PNS. Articulatory performance has most commonly been assessed with reference to lip (L) and jaw (J) movement, as these articulators are the most accessible to observation. Early investigations into the relationship between phonatory and articulatory co-ordination employed photoglottographic recordings in conjunction with acoustic recordings (Yoshioka & Löfqvist, 1981). Subsequently the use of electroglottographic (EGG) and electromyographic (EMG) data from the lower lip allowed the calculation of physiological response times (as opposed to acoustic response times), with PWS being found to have descriptively longer VOT than PNS (Bakker & Brutten, 1990). Further EMG studies have revealed a general pattern of greater displacement and greater variability in lip movements in PWS than in PNS. This pattern is also apparent in studies employing either a strain gauge or a light-tracking (IRED) approach, also measuring lower lip, upper lip and jaw (LL, UL, & J) displacement (cf. Table 1). When a strain gauge approach has been used to investigate the sequencing of speech motor movements (for UL, LL, J) it has been found that atypical sequencing may be a consequence of adaptations rather than a primary symptom (McClean, Kroll, & Loftus, 1990).

Ultrasound tongue imaging

Ultrasound, like EMA, captures kinematic information about the key active oral articulator, namely the tongue. Another aspect that sets UTI and EMA apart from studies that investigate only the external articulators such as lips and jaw is that the tongue is crucial for most consonants and all vowels. But even though the tongue plays a role in consonants and vowels alike, the sequencing and overlap in time and space of different parts of the tongue needs to be considered. UTI and EMA are

not identical however in their suitability for providing such data. When measuring the kinematics of the tongue, EMA typically offers a better temporal and 2D spatial resolution than UTI. There are two aspects however where UTI is advantageous over EMA, namely that it provides holistic midsagittal tongue surface data, and that its output is not limited to just three or four anterior data points. (Also, UTI is more accessible.) In terms of spatial resolution, UTI is equivalent to EMA in radial directions relative to the probe (sub millimetre accuracy), but is worse in circumferential measures, both as distance from the probe increases, and as the number of echopulse beams within a given field of view decreases (Wrench and Scobbie, 2011). Both techniques are poor at imaging the tongue tip, since EMA's coils interfere with articulation, while UTI loses its capacity to image the tip if it is masked by the jaw shadow or raised to create a sublingual air pocket.

Regarding the nature of the kinematic measures, they therefore draw on different underlying spatiotemporal data. While UTI provides images of almost the entire tongue surface moving in time and space in a two dimensional plane, EMA tracks the path of a few pre-determined flesh-points, typically but not necessarily in just two dimensions and just in the mid-sagittal plane. Typically for EMA three or four electromagnetic coils are glued on the anterior part of the tongue's upper surface as close to a midsagittal site based on the tongue's symmetrical morphology as possible, and nowadays coils are recorded as they move in 3D, with analysis based on a data reduction to 2D movement within a cranial midsagittal plane. Ultrasound instead samples movement of the tongue's surface through a single plane, and is typically orientated to cranial midsagittal orientation. It therefore captures an apparent mid-sagittal image of the tongue from near the tip right down to the root through space and time. This provides information not only about the tongue upper surface shape and location, but about tongue internal muscles (e.g. genioglossus), which can contribute to a principal components analysis. It is still considered sufficient in most research to consider only the wealth of surface data which both techniques provide, in apparent 2D motion, while remembering the different nature of these idealisations. Since the tongue's midline and the cranial midline need not correspond exactly at rest, and since they vary

during speech thanks to slight lateral asymmetries in speech production, the 2D data provided differ at source, even before we approach the holistic vs. fleshpoint differences. Finally, of course, other crucial lateral and constrictional aspects of spatiotemporal production ought to be considered for a full picture, which requires using other techniques, such as Electropalatography or MRI.

UTI is particularly relevant for clinical research, where we cannot know a priori where exactly to measure kinematics, for example, where the right place would be to place each EMA coil. The place of consonantal constriction may, for example, be more variable for experimental speakers with a speech disorder than for control speakers, and movement patterns of a coil in a suitable place for typical speech might be unrevealing for disordered speech. It is often not highlighted, in fact, that even for quantifying typical speech, the placing of an EMA coil is crucial, since slightly different coil placement provides a different kinematic trace, and different analytic values. Greater study of how variation in EMA coil placement affects kinematic measures is needed in order to ensure the validity of data and derived measures. The same is of course true of kinematic measures from ultrasound, as we will see.

UTI is more easily accessible and less invasive than EMA. This point is particularly relevant when recruiting and testing clinical populations of relatively low incidence (for example, at approximately 1% for stuttering, Craig, Hancock, Tran, Craig, & Peters, 2002), as UTI can be undertaken by a wider range of research teams and disciplines. The relatively non-invasive nature of UTI (compared to EMA) is valuable when working with populations who may be particularly sensitive to and atypical in their adaptations to alterations in sensorimotor feedback, since EMA requires that people speak with wires emerging from between the lips (though obviously some speakers may not tolerate the headset needed to stabilise the UTI probe). The great advantage of EMA however is that the data from each coil is perfectly suited for dynamic analysis, and there is large literature of established techniques (Schönle, Gäble, Wenig, Höhne, Schrader, & Conrad, 1987; Hoole & Nguyen, 1997). On the other hand, quantitative analysis of UTI is typically static, in terms of the shape of the tongue at a segmental target. It usually relates the ultrasound data to associated acoustic events

relative to which singular ultrasound frames are extracted (whether the acoustic midpoint, stop burst or maximum constriction). Static UTI analysis has been employed to explore articulation from a variety of angles. Video-based ultrasound, with data output rates of 30 frames per second (which can be deinterlaced to 60 fps if appropriate), can however also be used for timing analysis. For many purposes video rate output is as useful as high-speed ultrasound (Wrench and Scobbie, 2008), and it has been used to investigate socio-phonetic processes of timing (Lawson, Stuart-Smith, & Scobbie, 2014) and also processes of motor control (Zharkova, Hewlett, Hardcastle, & Lickley, 2014), including more specifically, coarticulation and intergestural timing (Gick & Campbell, 2003). Both experimental and theoretical evidence indicate that in order to investigate stuttering it is valuable to explore temporal as well as spatial aspects of speech execution. Despite the optimism of Wrench and Scobbie (2008), the low number of frames used in video ultrasound probably limits such kinematic analyses too much. Not only may the few frames that are available not be able to meaningfully capture the more subtle nature of articulatory movement, but more importantly, a slower scan rate at the probe combined with buffering of data to create the images results in both temporal smearing of the raw image, double tongues, and other spatial artefacts in the output images (Wrench and Scobbie, 2006). These make video data more suitable for analysis of the slow moving end points of articulatory-acoustic goals (i.e. the targets) than for kinematic analysis, especially of fast-moving articulations (Wrench and Scobbie, 2011). This is particularly important when investigating a disorder which essentially involves disruption to the smooth gestural flow of spoken output, where it is the process of articulatory-acoustic goal attainment which is or primary interest.

Dynamic analysis of ultrasound, to be similar to EMA, should therefore be based on a larger number of frames in the raw high-speed ultrasound data, for example, captured and stored at 120 frames per second or higher (Wrench and Scobbie, 2011). The large number of frames allows in-depth temporal and spatial investigation in principle. Articulatory events can be observed throughout the entire recording enabling the researcher to explore events that are less predictable and not acoustically salient. Both aspects about temporal and spatial resolution of UTI are beneficial

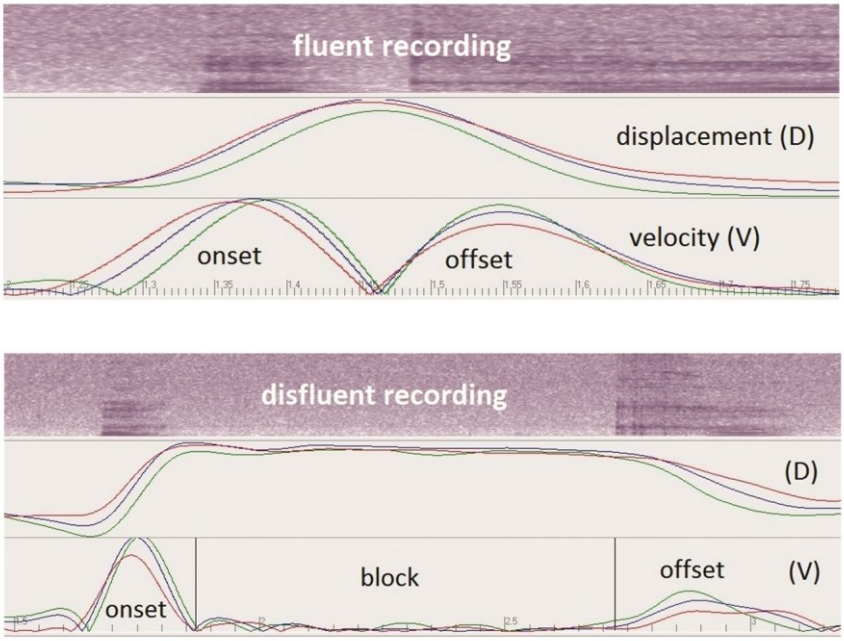
for detailed analysis of speech movements, even in qualitative analysis (Scobbie, Punnoose and Khattab, 2013). Similar to EMA, measures of duration and velocity can be obtained, to shed light on the trajectory of the tongue surface and its components. This has been useful in the investigation of degree of coarticulation (Zharkova et al., 2014) and inter-gestural timing movement (Strycharczuk and Scobbie, in press).

Fault-line hypothesis

The theoretical framework for the present paper is based on Wingate's Fault-Line hypothesis (Wingate 1988). The Fault-Line hypothesis responds to findings that PWS parse phrases based on syllables rather than utterances and that disfluencies typically commence on a consonant occurring on the first stressed syllable. Wingate claims that the main cause of disfluencies is the change in phonation (Wingate 1976), which leads him to hypothesise that PWS do not struggle to initiate the consonant (i.e. the syllable onset) or the following vowel (nucleus), but to transition between them. The Fault Line hypothesis (Wingate 1988) therefore postulates that disfluencies result from PWS struggling to formulate rhymes (nucleus and coda) especially in stressed syllables. The underlying cause according to Wingate lies in the difficulty of retrieval and encoding of syllables rhymes which delay or inhibit the integration of the onset with its rhyme. With reference to the Fault-Line hypothesis, dynamic analysis of ultrasound tongue imaging can be of avail in the analysis of stuttering data in that it allows quantifying and comparing lingual coordination (i.e. duration and velocity). Looking at the different movement patterns in fluent and disfluent speech (Figure 1) it may be that the closing consonantal gesture for the velar constriction is indeed similar in both cases, with the difference between them being attributable entirely to the long block, where the tongue body perseveres in palatal constriction. Assuming an underlying motor impairment in PWS, differences should be observable also in their apparently fluent speech.

Figure 1: Example displacement and velocity traces for a fluent (top) and a disfluent (bottom)

production of /ə kɑ/). Note that the disfluent production (1353ms) lasts approximately four times as long as the fluent production (327ms).



Aims

The purpose of the current study is to explore the potential of ultrasound tongue imaging as a tool to investigate motor coordination in the speech of PWS. In a pilot study we explore syllable initial coordination in the fluent speech of three people who stutter and compare it to that of three control speakers. Sample data from CV syllables where C corresponds to the velar consonant /k/ is examined in detail. From the range of possibilities, one specific measurement vector is identified as appropriate for the quantitative analysis of one dimensional movement, and the apparent speed of the tongue surface up and down this vector is investigated and presented. Measures of displacement and velocity were collected at the point of maximum displacement of the tongue surface. Holistic movements are subdivided into movement 'strokes', which are derived from directly observable kinematics, which are then interpreted in terms of the underlying gesture. Each stroke is defined as the period between two successive minima in movement velocity along the measurement vector (Tasko & Westbury, 2002). Two movement strokes are of particular interest

(Figure 1); the movement towards the consonantal constriction of /k/ (i.e. the onset to closure) and the movement away from the constriction into a steady state of the vowel (i.e. the offset). In fluent speech, each stroke has a ballistic character, with a single peak velocity, reflecting aspects of the underlying gestural control parameters. Stroke durations and their peak velocities are compared within and between groups for three different vowel contexts following the /k/.

The hypotheses are that coordination patterns in the fluent speech of PWS and PNS will be found to behave similarly in duration and/or peak velocity for movement onset, whereas they will differ for offset movements. This would signify no gestural difficulty when initiating the absolute syllable-initial consonant and moving into its constriction, but a problem in either gestural planning or implementation when transitioning from the consonant into the following vowel.

Materials and methods

Participants

Three experimental speakers and three control speakers are reported (cf. Table 2 for demographic information). Speakers all fulfilled the two main criteria of fitting into the stabilising headset and providing good ultrasound image quality. The group of experimental participants self-reported as having a persistent developmental stutter (i.e. a stutter with an onset by age eight) (Büchel & Sommer, 2004; Prasse & Kikano, 2008). Stuttering severity was assessed using both a formal assessment (SSI-IV) and an assessment of the individuals' experience of their stutter (OASES). Results from both the SSI-IV (Riley, 2009) and OASES (Yaruss & Quesal, 2006) classified stuttering severity ranging between mild-to-moderate and moderate-to-severe. For all speakers the last therapeutic intervention was a minimum of five years prior to recording. None of the speakers reported any lasting effect of the intervention. Speakers were recruited from the Edinburgh area. None of the participants reported any neurological, motor, auditory or visual impairment that could influence the outcome of the study. Speakers were compensated for their time with £15.

Table 2

Demographic information for speakers

	PWS 1	PNS 1	PWS 2	PNS 2	PWS 3	PNS 3
Gender	female	female	Male	Male	Male	Male
Age band	25-30	25-30	25-30	25-30	50-60	50-60
Handedness	Right	Right	Right	Right	Right	Right
Educational background	Post-graduate degree	Post-graduate degree	First Degree	First Degree	Information not provided	Post-graduate Degree
OASES	moderate (58/253)		mild-to-moderate (41/180)		moderate-to-severe (62/253)	
SSI-IV	moderate (26)		mild (19)		very severe (37)	

Stimuli

Data for the current study were part of a bigger corpus of recordings of adult speakers with a persistent developmental stutter. Target stimuli of the data presented are combinations of CV syllables with a voiceless velar stop (/k/) followed by a corner vowel (/i/ or /ɑ/) or schwa (/ə/). (It was decided that /u/ was too variable in placement in English to be used as a consistent context.) Each recording of a CV target item was preceded by a schwa (/ə/) to ensure a comparable lingual starting position. The preceding schwa was also useful in that it prevented bracing behaviours which would make it difficult to determine the time at which movement is initiated: the target word was therefore not in absolute utterance-initial position. Participants were instructed to stress the CV syllable following the schwa (i.e. to produce the pseudo noun phrases ‘a kaa’, ‘a kuh’, ‘a kee’). Fluent and disfluent recordings of the target stimuli /kɑ/, /kə/ and /ki/ will be presented separately.

Procedure

Participants were seated in front of a computer screen in a sound-treated recording room at Queen Margaret University. The ultrasound probe and a small microphone were attached to a stabilisation headset that participants wore during the recording session. The probe was oriented so

as to display a mid-sagittal configuration with the tongue tip to the right and the root of the tongue on the left (Figure 2). The headset was used to control and reduce movement of the ultrasound probe as well as to ensure clarity of the ultrasound image. The ultrasound PC and a second control PC connected by Ethernet were located in a neighbouring control room, and data capture, synchronisation and data storage were controlled with Articulate Assistant Advanced software v2.14 (Articulate Instruments Ltd, 2012). The researcher in the control room initiated the beginning and the end of each token recording. As soon as the recording was initiated by the researcher a fixation cross appeared on green background for 300ms. Following this 300ms delay, participants perceived a beep sound cueing them to read the prompt that appeared simultaneously on the screen. .The ultrasound machine that was used to record the data was an Ultrasonix SonixRP, which has the advantage of being particularly precise in timing of ultrasound data capture and storage and in audio-visual synchronisation (Wrench and Scobbie, 2008, 2011). Data was recorded at ~121 frames per second (fps) with 63 echopulse scan lines evenly spread over a 135 degree field of view (2.1° apart) . The maximum depth was set to 80mm and the echo return vectors had 412 samples resulting in a resolution of approximately 5 pixels per radial mm. The transducer frequency was 5MHz, capable of resolving at a radial resolution of approximately 1 mm (Articulate Instruments Ltd, 2012). The data is stored in scanline format, and reconstructed by AAA on the fly with radially interpolation to create a traditional fan-shaped image as the input to edge-fitting and subsequent analysis.

Fluency judgement

Recordings from PWS were categorised either as fluent or disfluent. The categorisation was necessary to compare perceptually fluent recordings from PWS with those of PNS employing quantitative measures. Disfluent data was identified in two steps: The first author inspected the visual ultrasound and audible acoustic data of the entire corpus and extracted data that appeared acoustically and/or articulatorily aberrant from repetitions of the same target stimulus produced elsewhere by the same speaker. The preselected ‘aberrant’ recordings together with ‘fluent’ control

versions of the same target stimulus were used for an objective evaluation. Twenty-five potentially disfluent recordings and 25 control recordings were randomised in an auditory judgement task. In a multiple forced choice experiment the recordings were presented in three randomised blocks (resulting in 150 stimuli overall) to five listeners. Listeners were trained linguists with no expertise in judging disfluent data. No hearing impairments were reported by listeners. After a brief introduction to the material, the recordings were presented to the listeners one by one and they were instructed to judge whether the material appeared to be fluent or disfluent. Listeners were also required to indicate how certain they were about each judgement on a 4 point scale. Recordings were regarded as clearly disfluent when at least four out of the five listeners rated the recording as 'disfluent' with certainty (3 or 4 points on the 4-point scale).

Analysis

Dynamic data was analysed in four steps: (1) acoustic landmarking of word and segment locations, (2) splining of the tongue contour, (3) determination of location for measurement vectors, and (4) kinematic annotations.

1. Acoustic landmarking

Acoustic data were exported from AAA into Praat (Boersma & Weenink, 2015) and semi-automatically annotated. An initial script distinguished silence from speech.[1] A following script opened each recording with the acoustic waveform and spectrogram. Automatically set boundaries were investigated and corrected when necessary. Additional boundaries were inserted distinguishing schwa, closure, release and vowel. Boundaries for vowels were based on periodic variation in the waveform preceding and following the voiceless consonant. The boundaries for the consonant were set at the onset of the stop consonant burst and at the onset of voicing for the

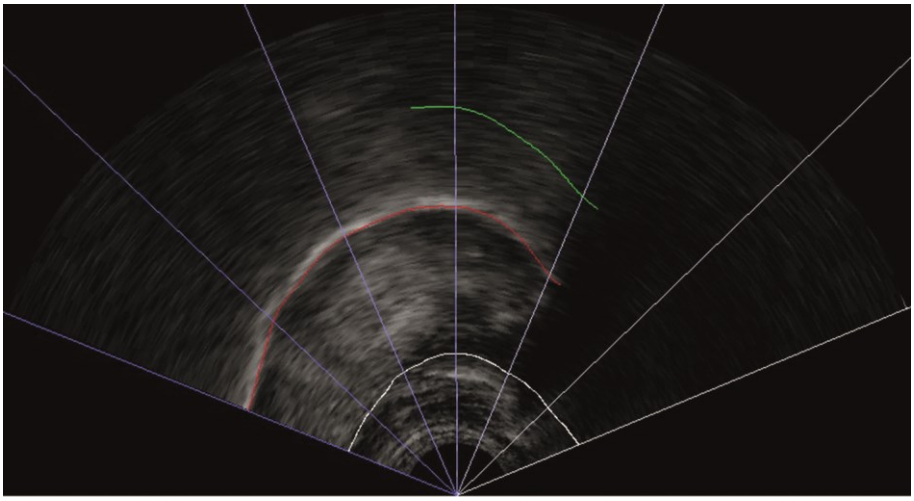
[1] The script settings were the following: Minimum pitch 60 Hz, Time steps: 0 s, silence threshold: -25 dB, minimum silent interval duration: 0.3 s, minimum sounding interval duration: 0.1 s.

following vowel. Acoustic landmark time-points were reimported into AAA and reintegrated with the audio / ultrasound data

2. Splining of the tongue contour

Splines were inserted to track the edge of the tongue contour. Splines are mathematical functions that are useful for fitting a smooth curve to data. A spline is fitted to the shape of the tongue by defining a number of points along the length of the tongue: in AAA a default fan-shaped grid provides 42 equally-spaced control points over the whole fan-shaped image. For the first-pass analysis reported here, splines were fitted to the data on a reduced temporal sample rate of 40 splines per second (i.e. on every third frame of actual data), mainly for logistical reasons of time.

Figure 2: Ultrasound image showing the mid-sagittal tongue configuration (with tongue tip to the right and tongue root to the left) with an overlaid spline (red line) framed by traces of the hard palate (green line) and the floor of the mouth (white line)

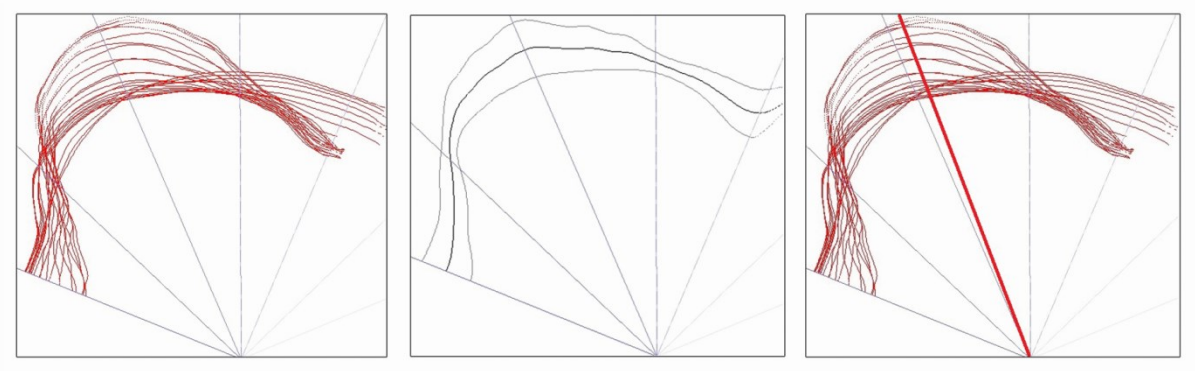


Semi-automated edge tracking was performed using AAA's built-in tracking functions on the frames displaying lingual movement from the schwa sound into the consonant /k/ and transitioning from /k/ onto the following vowel until the tongue contour reached a stable position following the release phase. An upper and a lower limit for automatic edge-fitting are specified for each frame in the same recording. To attach splines to each of the frames in a period of speech via the AAA tracking function, first an approximate tongue contour for the first frame is drawn manually. Its location is then refined semi-automatically with AAA's inbuilt 'snap-to-fit' function, a local search which scans along each of the 42 fan lines for the best dark-to-light edge. This function takes an input candidate spline and moves it to the clearest edge in its near neighbourhood, after which a built-in within-frame smoothing option can be applied to reduce radius-to-radius variation which can occur to the large number (up to 42) of closely-spaced knots (Articulate Instruments Ltd, 2012). This edge tracking spline-fitting function therefore identifies the location of the tongue surface and gives it a relative confidence rating indicating the validity of the data at each knot. It is possible to manually correct the spline by moving incorrect knots, but particularly useful is the facility to set confidence to 0% at the anterior and posterior edges of the tongue surface in the image, which makes these irrelevant parts of the spline invisible to the user and to subsequent analysis.

3. Measurement vector

The fitted first spline described above is the starting point for an automated edge tracking of the tongue surface contour throughout the subsequent frames of the recording. The tracking function bases a new spline in a frame on the shape already finalised for the preceding frame, then does a snap-to-fit local search for the new edge. The AAA tracking function re-iterates the snap-to-fit function automatically through a series of frames. Local search edge tracking works well for tracking dynamic changes in the mid-sagittal curve given a suitably high frame rate and clear images. These require a high underlying probe scan rate captured digitally because each frame is both a clear snapshot of a short time interval and only slightly different to the one preceding it.

Figure 3: Superimposed splines in 2D (left); confidence and mean standard deviations (middle); measurement vector at selected fan line in velar area at maximum lingual displacement (right)

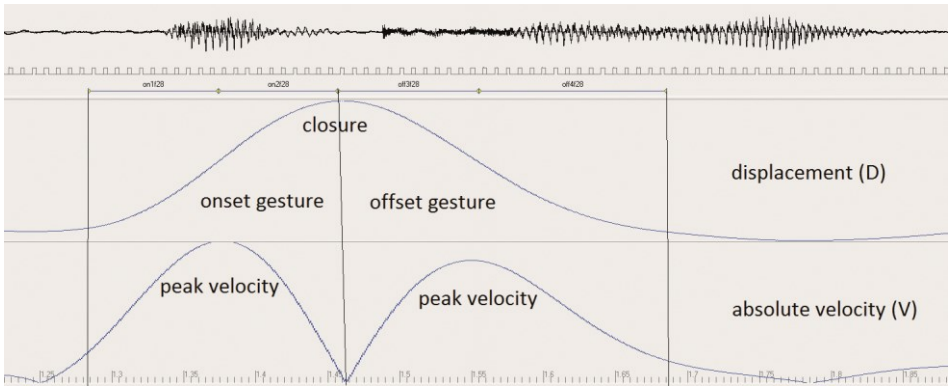


Tracking the splines throughout the frames is based on a combination of edge detection and brightness detection. Tracking was interrupted and new manual starting point for the splines was defined manually if artefacts in the ultrasound image led the automated tracking to go astray. To determine a suitable measurement vector for kinematic analysis, the splines that display onset, closure and offset of the tongue movement can be superimposed, to create a 2D image that informs about the extent of tongue movement at different points in the vocal tract, indirectly reflecting differential movement extent of different areas on the tongue surface (Iskarous, 2004). Splines are superimposed separately for each speaker and context because each vector needs to be not just speaker-specific but specific to the comparisons being made. The resulting image (Figure 3 a) is used to estimate the strength of the signal as well as to establish areas on the tongue surface where displacement is largest. For the example given, investigating a velar consonant, the measurement vector should a priori be placed in the velar area: indeed this was where lingual displacement was largest. For all measurements presented here, the candidate vectors were all fan radii, given the instrumentally high resolution of ultrasound in radial directions and the nature of constrictions for /k/ in the vocal tract relative to the probe, though note that other orthogonal vectors could be used

if thought necessary. The specific vector for analysis was chosen on objective criteria as follows. A mean spline was created based on the mean values for each spline-knot at each of the 42 fan radii, from the dynamic articulatory event of interest, with standard deviation and confidence indicated. All subsequent measurements were taken along the scan line with the greatest standard deviation (Figure 3 b), which was taken as indicative of the area of greatest movement. A relatively high confidence of the semi-automatic AAA spline fitting (at least 85% overall) was used as a threshold for the validity of data. In different conditions, the radial line for which the lingual movement was largest (i.e. largest value of standard deviation from the mean spline) was adopted as the measurement vectors for kinematic analysis of tongue surface speed (cf. red line in Figure 3 c).

4. Annotations

Figure 4: Displacement and velocity traces with labels indicating onset and offset movement 'strokes' with the relative peak velocity



Both displacement and velocity were calculated by AAA for the spline as it changed location along the measurement vector. Displacement measures (D) indicate the radial distance from the origin of where a spline crosses the vector (Figure 4). Displacement and absolute velocity (V) data were captured for the relevant area including movement onset, lingual closure and the offset movement away from palatal constriction until a stable position for the vowel was reached. Based on the absolute velocity two gestural 'strokes' could typically be readily identified for the production of the CV target stimulus. An inbuilt 'find function' (Articulate Instruments Ltd, 2012) was used to semi-automatically create annotations based on the absolute velocity profile (lower tier) starting

from the point of zero velocity at acoustic closure. This point of zero velocity is preceded and followed by increases in velocity signifying the movement towards and away from consonantal constriction. Two regions were therefore identified based on the velocity trace for each recorded CV syllable, reaching back and forward in time from the stable target.

1. onset: from movement initiation away from a relatively stable position, via maximum velocity, up to consonantal constriction where absolute velocity reaches zero, and
2. offset: from consonantal constriction with zero absolute velocity, via maximum velocity, until the tongue reaches a stable position for the vowel.

Measures for the beginning of the movement onset phase and the end of the offset phase are more arbitrary, ambiguous and sensitive to subtle but irrelevant articulatory movement, so were fixed at a timepoint where the velocity exceeded a threshold of 20% of the peak velocity when moving towards / away from the closure for the velar stop. This is a typical procedure in EMA studies (Pouplier & Waltl, 2008; Tasko & Westbury, 2002), and avoids undesirable and theoretically misleading variation in duration measurement

Results

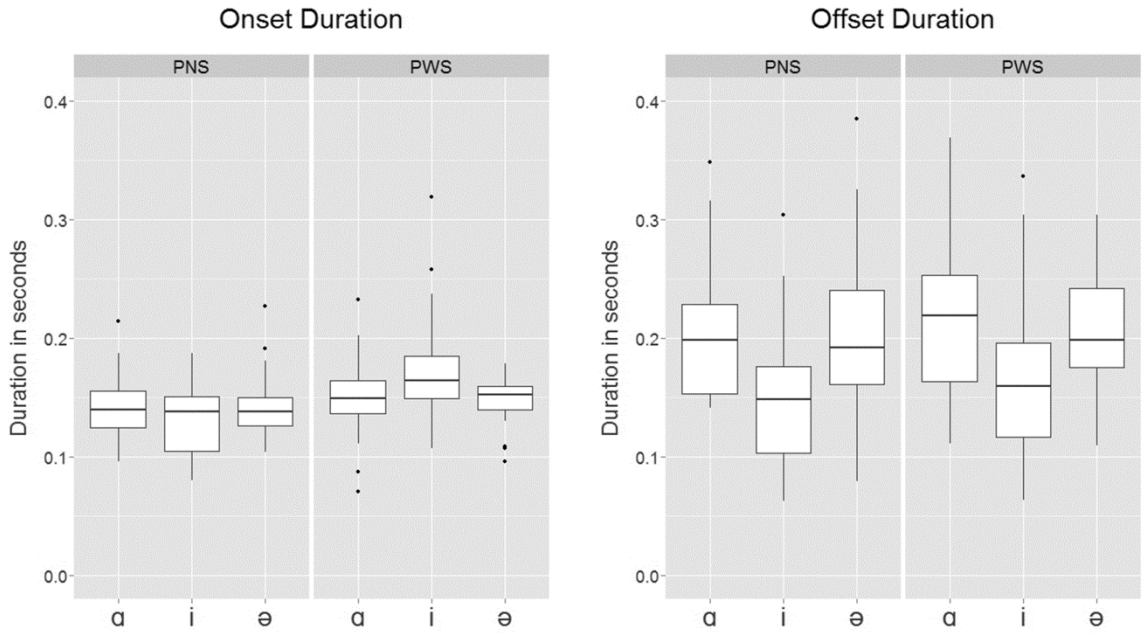
For reference we include the results of statistical analyses on this pilot data set, but we caution that these analyses are most likely underpowered, particularly where speaker group difference are concerned (power analyses indicate a minimum requirement of 8 participants per group in order to achieve a ß -level of 0.08 at an α-level of 0.05). Results are reported where (i) they concern group differences and/or, (ii) they reveal significant differences. The stability/variability of articulatory coordination was investigated in the fluent speech of PWS and PNS, which are shown in the graphs below. Only a few recordings of disfluent productions by PWS were available. The perceptual analysis resulted in 8.8 % of PWS recordings (8 out of 91) being identified as clearly disfluent. Their comparison to the fluent data is secondary to the comparison of fluent speech

between groups. Three distinct steps were undertaken: (1) Onset and offset movement durations across multiple repetitions of the syllable initial consonant /k/ were compared across vowel context for both groups. (2) Maximum velocities for onset and offset movements were examined again contrasting vowel context across the two groups of PWS and PNS. In order to verify differences in the lingual movement of PWS and control speakers, descriptive statistics were calculated for the durations and peak velocities for both groups. Median values together with standard deviation are reported for 195 recordings (total of 203 less 8 disfluent recordings) from 6 speakers (3 PWS and 3 PNS) distributed over three vowel contexts (i.e. /kɑ/ (n=65), /ki/ (n=66) and /kə/ (n=64). Each recording comprises an onset and an offset region (see Figure 4).

Duration measures

Durations for onset and offset were calculated and are displayed in the two graphs representing the movement towards (Figure 5 a) the velar /k/ closure and away from (Figure 5 b) consonantal constriction for both groups. Data is presented for the three different vowel contexts (i.e. /ɑ/, /i/, /ə/) shown in the panels of the graphs.

Figure 5: Duration measures for (a) onset and (b) offset by group (PNS vs. PWS) and vowel context (/ɑ, i, ə/)



Independent of vowel context and group, offsets are on average 0.04 s longer than onsets. While the duration for PWS increases from 0.16 s to 0.20 s, the duration for PNS increases from 0.14 s to 0.18 s. When the data are modelled, only slight duration differences between groups can be observed (onsets: ß=0.013 (se=0.009), t=1.43; offsets: ß=0.016 (se=0.014), t=1.225). Offset duration further shows increased variation when compared to onset duration with standard deviation increasing from approximately 0.03 up to 0.07 s. Durations for both onset and offset phases also tend to behave comparably across the three vowel contexts. Only the high vowel /i/ appears to affect offset movement duration, noticeably reducing it (/kɑ/ vs. /ki/: ß=-0.046 (se=0.019), t=-2.453; /kɑ/ vs. /kə/: ß=-0.001(se=0.013), t=-0.110[2]; cf. Table 3 for descriptive data). While overall offset durations for both vowel contexts /ɑ/ and /ə/ are fairly constant (PWS: M

[2] Including the interaction between prompt type and speaker group did not improve model fit.

0.21/0.20 s; PNS: M 0.20 s/0.20 s), a decrease in offset duration (PWS: M 0.17 s; PNS: M 0.15 s) for /i/ vowel context is apparent.

Table 3

Duration measures for onset and offset strokes by vowel context and speaker group

				/a/		/uh/		/i/	
		Onset	Offset	Onset	Offset	Onset	Offset	Onset	Offset
All	Mean (SD)	0.15 (0.03)	0.19 (0.07)	0.15 (0.03)	0.21 (0.06)	0.14 (0.02)	0.20 (0.07)	0.15 (0.04)	0.16 (0.06)
PWS	Mean (SD)	0.16 (0.04)	0.20 (0.07)	0.15 (0.03)	0.21 (0.06)	0.15 (0.02)	0.20 (0.05)	0.17 (0.04)	0.17 (0.07)
PNS	Mean (SD)	0.14 (0.03)	0.18 (0.07)	0.14 (0.02)	0.20 (0.05)	0.14 (0.02)	0.20 (0.07)	0.13 (0.03)	0.15 (0.06)
PWS (disfluent)		0.22 (0.08)	0.24 (0.11)						

Tabel notes

Looking at data from only PWS and comparing the duration data for fluent (n of tokens=83) with that of disfluent (n of tokens=8) recordings, there is a noticeable difference in duration for onsets (fluent: M 0.16 s; disfluent: M 0.22 s) as well as offsets (fluent: M 0.20 s; disfluent: M 0.24 s). In the disfluent data we find a similar pattern to that seen in the fluent data; shorter and less variable onsets compared to offsets (onset: M 0.22 ms, SD 0.08 s; offset: M 0.24 s, SD 0.11 s).

Peak velocity measures

Peak velocities for the tongue approaching the palate (i.e., onset velocity; Figure 6 a) are comparable between groups (PWS: M 124 mm/s, SD = 52; PNS: M 126 mm/s, SD = 32 mm/s). In both speaker groups, mean peak velocity for offset movements (PWS: M 79 mm/s, SD = 35 mm/s; PNS: M 104 mms/s, SD = 54 mm/s) is lower than for onset movements. This difference is stronger in PWS (Figure 6b). Statistically, group difference is not significant for either onsets (ß=13.39 (se=34.38), t=0.390) or offsets (ß=14.06 (se=31.52), t=0.446). However, it may be of theoretical significance that,

as a group, PNS display greater variability in offset velocity than onset velocity, whereas for PWS the reverse is true.

Figure 6: Peak velocity measures for (a) onset and (b) offset by group (PNS vs. PWS) and vowel context (/ɑ, i, ə/)

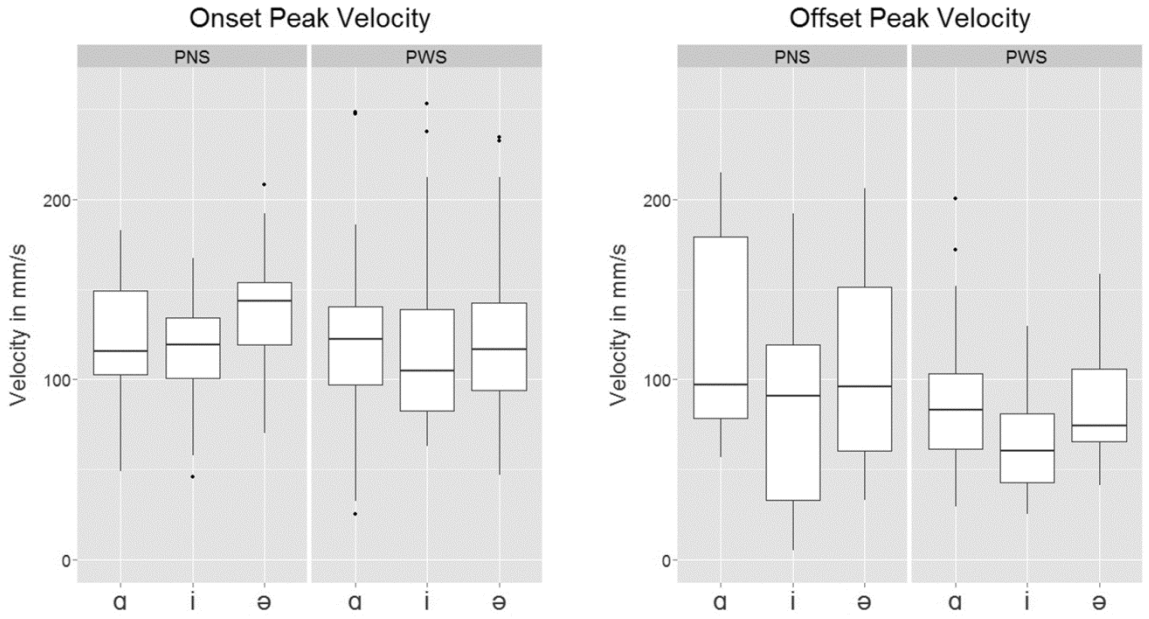


Table 4

Peak velocity measures for onset and offset strokes by vowel context and speaker group

				/a/		/uh/		/i/	
		Onset	Offset	Onset	Offset	Onset	Offset	Onset	Offset
All	Mean	125.08	93.83	122.50	109.38	133.25	97.59	119.70	74.87
	(SD)	(41.85)	(48.36)	(41.78)	(51.46)	(40.96)	(43.49)	(42.18)	(43.82)
PWS	Mean	123.72	79.34	121.91	89.69	125.95	87.05	123.29	63.23
	(SD)	(52.30)	(34.72)	(53.02)	(40.62)	(51.66)	(31.05)	(53.94)	(26.72)
PNS	Mean	126.09	104.40	122.90	122.00	138.57	105.50	116.70	84.24
	(SD)	(32.21)	(54.01)	(33.02)	(54.13)	(30.66)	(49.84)	(29.55)	(52.33)

Vowel context affects peak velocity for both groups (cf. Table 4). The vowel dependent disparity in mean peak velocity is most evident for offset movements (/kɑ/ vs. /ki/: ß=-34.50

(se=17.95), t=-1.921; /kɑ/ vs. /kə/: ß=-12.04 (se=5.55), t=--2.169)[3], particularly when produced by PWS (with M 63 mm/s for /i/ compared to M 89 mm/s for /ɑ/ and M 87 mm/s for /ə/). Differences were also observable for onset velocity for /kɑ/ vs. /kə/ (ß=11.103 (se=5.401), t=2.056) but not for /kɑ/ vs. /ki/ (ß=0.437 (se=9.074), t=0.048).

In summary, offset (compared to onset) peak velocities are lower for PWS than PNS and display a larger vowel effect. For duration measures onset (when compared to offset) durations are shorter for both groups with overall slightly lower values for PNS than PWS (fluent < disfluent recordings) (cf. Table 3). Patterns are consistent across vowel contexts with the exception of /i/.

Discussion

A novel approach to analysing dynamic ultrasound data was presented, and applied to the fluent speech of speakers with a history of fluency problems (PWS). As predicted, PWS and controls (PNS) did not differ on duration or peak velocity measures when approaching an initial consonantal constriction (onset). Further, no statistically significant difference between groups was found for offsets. From this perspective, all speakers were equally fluent. The non-significance for between-group testing could result from (1) the nature of the stimuli or (2) the low number of speakers. Having included measures from only 'fluent' recordings, we by definition have eliminated the more apparent differences that could be expected otherwise (i.e., in the coordination of movements between fluent and disfluent speech). The low number of speakers makes larger values for variability more likely, reducing power to detect significant results. Results are therefore meaningful for descriptive analysis more so than for inferential statistics.

Vowel effects were observed for offsets affecting both duration and peak velocity measures. Effects observed for /i/ on offset duration measures may be due to an increased coarticulatory effect of /i/ onto /k/, decreasing the trajectory through the fronting of /k/. The vowel effects observed for offset peak velocity measures, on the other hand, may be due to an enhanced looping

[3] Again, including the interaction between prompt type and speaker group did not improve model fit.

effect on /i/ in combination with /k/ (cf. Mooshammer, Hoole, & Kühnert, 1995). In both cases forward movement would intersect the measurement vector (i.e. forward movement is essentially perpendicular to the measurement vector): it would not be measureable. However, it would most probably reduce the movement measured on the vector in these results. The more general vowel effects observed for velocity offsets further indicate the degree of accuracy of the proposed method.

Referring to descriptive data, results provide support for the notion that PWS do not struggle when moving towards the consonant, but do when transitioning from the consonant into the vowel. Both groups appear to use different strategies while still reaching the vowel target at the same time. Differences were observable for peak velocity measures in offsets, which are the transitions away from consonantal constriction towards a stable position in the following vowel. The descriptive statistics presented strongly suggest that PWS and PNS differed regarding both mean peak velocity and its degree of variation. Measures for PWS displayed observably lower means for peak velocity compared to those from control speakers. Control speakers on the other hand showed an overall larger variability of peak velocity. The lower overall peak velocity in offsets could suggest that PWS have fundamentally lower acceleration/deceleration compared to PNS.

Because experimental speakers were adults who have had their stutter since childhood, they are highly likely to have found strategies to maintain perceptually fluent speech, and the generally lower peak velocity could be just one. Despite the different histories of stuttering therapy, the shared strategy of lower peak velocity during release could reflect its efficacy to the user in overcoming struggles in maintaining fluency. Further data is required to verify the present results.

In accordance with Wingate's Fault-Line Hypothesis, these preliminary results show that PWS do not struggle when initiating the syllable initial consonant. In contrast, the differences between groups in peak velocity may indicate struggles PWS have when transitioning on to the following vowel. The differences could demonstrate difficulty forming and integrating syllable rhymes with their onsets. The fact that kinematic differences between groups show in apparently

fluent speech could be an indicator for an underlying motor control impairment not limited to temporary disruptions in the speech flow.

Because differences between groups only affect velocity, but not duration measures, they could only be observed when looking at articulatory data. The approach we presented for kinematic ultrasound analysis relies crucially on an ability to observe dynamic movement of the tongue, a universal articulator, rather than just the lips or jaw. It allowed us to closely investigate the movement of the tongue towards and away from a consonantal constriction. The dynamic nature of the data made the trajectory of the tongue surface clearly observable and articulatory events could be analysed based on kinematics. In our approach we have attempted to create a systematic method of ultrasound kinematics that (1) is replicable, (2) allows for inter-speaker comparison, (3) is modifiable for different places of articulation and (4) can be simply extended, e.g. to other non-radial vectors. Movement trajectories were broken down into 'strokes', which refer to transitions between two articulatory gestures. Breaking down trajectories into 'stroke' duration and 'stroke' peak velocity guaranteed a more in-depth examination eventually revealing kinematic differences between groups. The approach presented uses maximum displacement as a referent for measures of duration and peak velocity. Starting from kinematic movement patterns, measures respond and adapt to the individuals' articulatory setting and tongue surface trajectory. This aspect renders inter-speaker comparison legitimate. Moreover, because measures are based on maximum displacement relative to the place of articulation of a sound, our approach is not limited to a specific place of articulation, but can also be adapted for a variety of sounds with differing places of articulation.

Concluding remarks

Ultrasound tongue imaging is an easily accessible instrument that is non-invasive and provides relatively high quality images. Notwithstanding, a number of impediments need to be considered and overcome. Owing to the fact that ultrasound, particularly in the study of dynamic data, is a fairly new in the research of speech movements there is a lack of established analytical procedures tested in different laboratories from practical and statistical perspectives. Moreover,

traditional UTI analysis relies heavily on 'eye-balling' in order to determine the area of interest on the tongue surface. 'Eye-balling' however requires extensive experience with data of that particular kind. A more replicable/reproducible approach to methodological improvement was set out and with further enhancements this sort of approach could help motivate the development of ultrasound kinematics by the wider research community. To validate the proposed method further testing should be conducted using larger data sets also including, for example, alveolar /t/ or fricative /s/ targets. Also, to test the reliability of the proposed method, data should be analysed across sessions of the same speaker. It is to be hoped that simultaneous vector-based UTI and fleshpoint-based EMA or replicated UTI/EMA datasets will test and verify kinematic ultrasound.

While we have seen that some kinematic UTI analysis can be done in the style of (or at least inspired by) well-established methods for the analysis of EMA data (i.e. adaptation of definitions of units of measure such as movement 'strokes'), others need to be defined anew. One of the 'to-be-defined' aspects bears on the general lack of a common referent that measures could be related to. This comes for free, if somewhat arbitrarily, with EMA, since a coils is glued to the articulators and not removed within a session. Ultrasound however provides a wealth of possible vectors for measurement within a session, and it is not clear how to solve the data-reduction problem. But both techniques face similar challenges when we consider how best to orientate and relate data across sessions and across participants. For ultrasound, we need to define an angle or referent line that is shared across speakers. For EMA, we need to define coil placements that are both optimal and shared.

While our approach was sensitive to different vowel contexts, it clearly needs to be run on larger amounts of data to be able to quantify its sensitivity to different consonantal targets and vowel contexts. We predict that it will be necessary to make use of non-radial measurement vectors, for example. The difficulty lies in defining how to locate them in order to make data available for inter-speaker comparison. In sum, we presented an approach in which we defined the location of

maximum radial displacement relative to the probe centre, so that the maximal dynamic variation could be used as a defining characteristic, across all speakers, for a common kinematic analysis.

Acknowledgements

We thank all participants who gave up their time to complete the experiment. Special thanks go to Alan Wrench, Patrycja Strycharczuk, Steve Cowen, Ian Finlayson, and Jana Walzog for their constant support. We gratefully acknowledge the support of Queen Margaret University, particularly the Clinical Audiology and Speech and Language Research Centre.

References

Adams, D. C. (1999). Methods for shape analysis of landmark data from articulated structures. *Evolutionary Ecology Research, 1*(8), 959-970.

Articulate Instruments Ltd. (2012). *Articulate assistant user guide: Version 1.18. edinburgh*. Edinburgh, UK: Articulate Instruments Ltd.

Bakker, K., & Brutten, G. J. (1990). Speech-related reaction times of stutterers and nonstutterers diagnostic implications. *Journal of Speech and Hearing Disorders, 55*(2), 295-299.

Boersma, P., & Weenink, D. (2015). *Praat: Doing phonetics by computer* (Version 5.4.05, retrieved 17 February 2015 ed.). http://www.praat.org/

Brocklehurst, P. H., & Corley, M. (2011). Investigating the inner speech of people who stutter: Evidence for (and against) the covert repair hypothesis. *Journal of Communication Disorders, 44*(2), 246-260. doi:10.1016/j.jcomdis.2010.11.004

Büchel, C., & Sommer, M. (2004). What causes stuttering? *PLoS Biology, 2*(2), 159-163.

Caruso, A. J., Abbs, J. H., & Gracco, V. L. (1988). Kinematic analysis of multiple movement coordination during speech in stutterers. *Brain : A Journal of Neurology, 111 (Pt 2)*(Pt 2), 439-456.

Chang, S., Ohde, R. N., & Conture, E. G. (2002). Coarticulation and formant transition rate in young children who stutter. *Journal of Speech, Language and Hearing Research, 45*(4), 676-688.

Craig, A., Hancock, K., Tran, Y., Craig, M., & Peters, K. (2002). Epidemiology of stuttering in the community across the entire life span. *Journal of Speech, Language, and Hearing Research, 45*(6), 1097-1105.

Cross, D. E., & Luper, H. L. (1979). Voice reaction time of stuttering and nonstuttering children and adults. *Journal of Fluency Disorders, 4*(1), 59-77.

De Nil, L. F., & Brutten, G. (1991). Voice onset times of stuttering and nonstuttering children: The influence of externally and linguistically imposed time pressure. *Journal of Fluency Disorders, 16*(2), 143-158.

Di Simoni, F. G. (1974). Preliminary study of certain timing relationships in the speech of stutterers. *The Journal of the Acoustical Society of America, 56*(2), 695-696.

Gick, B., & Campbell, F. (2003). Intergestural timing in english /r/. *Proceedings of the 15th International Congress of Phonetic Sciences,* 1-4.

Harbison Jr, D. C., Porter Jr, R. J., & Tobey, E. A. (1989). Shadowed and simple reaction times in stutterers and nonstutterers. *The Journal of the Acoustical Society of America, 86*(4), 1277-1284.

Healey, E. C., & Ramig, P. R. (1986). Acoustic measures of stutterers' and nonstutterers' fluency in two speech contexts. *Journal of Speech, Language, and Hearing Research, 29*(3), 325-331.

Hoole, P., & Nguyen, N. (1997). Electromagnetic articulography in coarticulation research. *Forschungsberichte Des Instituts Für Phonetik Und Sprachliche Kommunikation Der Universität München, 35*, 177-184.

Horii, Y. (1984). Phonatory initiation, termination, and vocal frequency change reaction times of stutterers. *Journal of Fluency Disorders, 9*(2), 115-124.

Iskarous, K. (2005). Patterns of tongue movement. *Journal of Phonetics, 33*, 363-381.

Kleinow, J., & Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language and Hearing Research, 43*(2), 548-559.

Lawson, E., Stuart-Smith, J., & Scobbie, J. M. (2014). A mimicry study of adaptation towards socially-salient tongue shape variants. *University of Pennsylvania Working Papers in Linguistics, 20*(2), 99-110.

Max, L., Caruso, A. J., & Gracco, V. L. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *Journal of Speech, Language, and Hearing Research, 46*(1), 215-232.

Max, L., & Gracco, V. L. (2005). Coordination of oral and laryngeal movements in the perceptually fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research, 48*(3), 524-542.

McClean, M. D., Kroll, R. M., & Loftus, N. S. (1990). Kinematic analysis of lip closure in stutterers' fluent speech. *Journal of Speech, Language, and Hearing Research, 33*(4), 755-760.

McClean, M. D., Tasko, S. M., & Runyan, C. M. (2004). Orofacial movements associated with fluent speech in persons who stutter. *Journal of Speech, Language, and Hearing Research, 47*(2), 294-303.

Namasivayam, A. K., & van Lieshout, P. (2008). Investigating speech motor practice and learning in people who stutter. *Journal of Fluency Disorders, 33*(1), 32-51.

Namasivayam, A. K., & van Lieshout, P. (2011). Speech motor skill and stuttering. *Journal of Motor Behavior, 43*(6), 477-489.

Pouplier, M., & Waltl, S. (2008). Articulatory timing of coproduced gestures and its implications for models of speech production. *Proceedings of the 8th International Seminar on Speech Production,* 19-22.

Prasse, J. E., & Kikano, G. E. (2008). Stuttering: An overview. *American Family Physician, 77*(9), 1271-1276.

Riley, G. D. (2009). *SSI-4 stuttering severity instrument*.

Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., & Conrad, B. (1987). Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language, 31*(1), 26-35.

Scobbie, J. M., Punnoose, R., & Khattab, G. (2013). Articulating five liquids: a single speaker ultrasound study of Malayalam. In L. Spreafico and A. Vietti (eds) *Rhotics: New Data and Perspectives*. Bozen-Bolzano: BU Press. 99-124.

Smith, A., Sadagopan, N., Walsh, B., & Weber-Fox, C. (2010). Increasing phonological complexity reveals heightened instability in inter-articulatory coordination in adults who stutter. *Journal of Fluency Disorders, 35*(1), 1-18.

Starkweather, C. W., & Myers, M. (1979). Duration of subsegments within the intervocalic interval in stutterers and nonstutterers. *Journal of Fluency Disorders, 4*(3), 205-214.

Strycharczuk, P., & Scobbie, J. M. (in press). Velocity measures in ultrasound data: gestural timing of post-vocalic /l/ in English. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS),* Glasgow, UK.

Tasko, S. M., & Westbury, J. R. (2002). Defining and measuring speech movement events. *Journal of Speech, Language, and Hearing Research, 45*(1), 127-142.

Van Riper, C. (1982). *The nature of stuttering* Prentice Hall.

Watson, B. C., & Alfonso, P. J. (1982). A comparison of LRT and VOT values between stutterers and nonstutterers. *Journal of Fluency Disorders, 7*(2), 219-241.

Wingate, M. E. (1976). *Stuttering: Theory and treatment.* Irvington.

Wingate, M. E. (1988). *The structure of stuttering: A psycholinguistic analysis* Springer Verlag.

Wrench, A. A., & Scobbie, J. M. (2006). Spatio-temporal inaccuracies of video-based ultrasound images of the tongue. *Proceedings of the 7th International Seminar on Speech Production,* 451-458.

Wrench, A. A., & Scobbie, J. M. (2008). High-speed cineloop ultrasound vs. video ultrasound tongue imaging: Comparison of front and back lingual gesture location and relative timing. *Proceedings of the Eighth International Seminar on Speech Production (ISSP),* 57-60.

Wrench, A. A., & Scobbie, J. M. (2011). Very high frame rate ultrasound tongue imaging. *Proceedings of the 9th International Seminar on Speech Production (ISSP),* 155-162.

Yaruss, J. S., & Quesal, R. W. (2006). Overall assessment of the speaker's experience of stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *Journal of Fluency Disorders, 31(2),* 90–115.

Yoshioka, H., & Löfqvist, A. (1981). Laryngeal involvement in stuttering. *Folia Phoniatrica Et Logopaedica, 33*(6), 348-357.

Zharkova, N., Hewlett, N., Hardcastle, W. J., & Lickley, R. J. (2014). Spatial and temporal lingual coarticulation and motor control in preadolescents. *Journal of Speech, Language, and Hearing Research, 57*(2), 374-388.

Zimmermann, G. (1980). Articulatory dynamics of fluent utterances of stutterers and nonstutterers. *Journal of Speech, Language, and Hearing Research, 23*(1), 95-107.